

CSC 261/461

Database Systems

Eustrat Zhupa

October 3, 2018



UNIVERSITY of
ROCHESTER

Database Design - Informal Guidelines

Issues

A bad database design may cause:

- ▶ Update anomalies
- ▶ Insertion anomalies
- ▶ Deletion anomalies



Example

EMPLOYEE

Ename	Ssn	Bdate	Address	Dnumber
Smith, John B.	123456789	1965-01-09	731 Fondren, Houston, TX	5
Wong, Franklin T.	333445555	1955-12-08	638 Voss, Houston, TX	5
Zelaya, Alicia J.	999887777	1968-07-19	3321 Castle, Spring, TX	4
Wallace, Jennifer S.	987654321	1941-06-20	291 Berry, Bellaire, TX	4
Narayan, Ramesh K.	666884444	1962-09-15	975 Fire Oak, Humble, TX	5
English, Joyce A.	453453453	1972-07-31	5631 Rice, Houston, TX	5
Jabbar, Ahmad V.	987987987	1969-03-29	980 Dallas, Houston, TX	4
Borg, James E.	888665555	1937-11-10	450 Stone, Houston, TX	1

DEPARTMENT

Dname	Dnumber	Dmgr_ssn
Research	5	333445555
Administration	4	987654321
Headquarters	1	888665555

DEPT_LOCATIONS

Dnumber	Dlocation
1	Houston
4	Stafford
5	Bellaire
5	Sugarland
5	Houston

Example

Redundancy						
EMP_DEPT						
Ename	Ssn	Bdate	Address	Dnumber	Dname	Dmgr_ssn
Smith, John B.	123456789	1965-01-09	731 Fondren, Houston, TX	5	Research	333445555
Wong, Franklin T.	333445555	1955-12-08	638 Voss, Houston, TX	5	Research	333445555
Zelaya, Alicia J.	999887777	1968-07-19	3321 Castle, Spring, TX	4	Administration	987654321
Wallace, Jennifer S.	987654321	1941-06-20	291 Berry, Bellaire, TX	4	Administration	987654321
Narayan, Ramesh K.	666884444	1962-09-15	975 FireOak, Humble, TX	5	Research	333445555
English, Joyce A.	453453453	1972-07-31	5631 Rice, Houston, TX	5	Research	333445555
Jabbar, Ahmad V.	987987987	1969-03-29	980 Dallas, Houston, TX	4	Administration	987654321
Borg, James E.	888665555	1937-11-10	450 Stone, Houston, TX	1	Headquarters	888665555



Database Design - Informal Guidelines

Guideline 2

- ▶ Design a schema that does not suffer from the insertion, deletion and update anomalies.
- ▶ If there are any anomalies present, then note them so that applications can be made to take them into account.



Functional dependencies

FDs

- ▶ Used to specify formal measures of the "goodness" of relational designs
- ▶ Are constraints that are derived from the meaning and interrelationships of the data attributes
- ▶ A set of attributes X **functionally determines** a set of attributes Y if the value of X determines a unique value for Y

[2]



UNIVERSITY of
ROCHESTER

Functional dependencies

Definition

- ▶ $X \rightarrow Y$ holds if whenever two tuples have the same value for X , they must have the same value for Y .
- ▶ For any two tuples t_1 and t_2 in any relation instance $r(R)$: If $t_1[X] = t_2[X]$, then $t_1[Y] = t_2[Y]$
- ▶ $X \rightarrow Y$ in R specifies a constraint on all relation instances $r(R)$
- ▶ FDs are derived from the real-world constraints on the attributes

[2]



Functional dependencies

Example 1

EMP_PROJ

Ssn

Pnumber

Hours

Ename

Pname

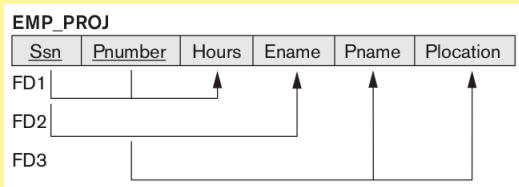
Plocation



UNIVERSITY of
ROCHESTER

Functional dependencies

Example 1



- ▶ $Ssn \rightarrow Ename$
- ▶ $Pnumber \rightarrow \{Pname, Plocation\}$
- ▶ $\{Ssn, Pnumber\} \rightarrow Hours$



Functional dependencies

Example 2

TEACH

Teacher	Course	Text
Smith	Data Structures	Bartram
Smith	Data Management	Martin
Hall	Compilers	Hoffman
Brown	Data Structures	Horowitz

[3]



UNIVERSITY of
ROCHESTER

Functional Dependencies

Functional Dependency

- ▶ An FD is a property of the attributes in the schema R
- ▶ The constraint must hold on every relation instance $r(R)$
- ▶ If K is a key of R , then K functionally determines all attributes in R
 - ▶ we never have two distinct tuples with $t_1[K] = t_2[K]$



Functional Dependencies

- ▶ In order to define the FDs, we need to understand the meaning of the attributes involved and the relationship between them.
- ▶ An FD is a *property* of the attributes in the schema R
 - ▶ Given the instance of a relation, all we can conclude is that an FD *may* exist between certain attributes.
 - ▶ What we can definitely conclude is that certain FDs *do not exist* if there are tuples that violate the dependencies.



Functional Dependencies

- ▶ A relation $R(A, B, C, D)$ with its extension is shown below.
- ▶ Which FDs may exist in this relation?

A	B	C	D
a1	b1	c1	d1
a1	b2	c2	d2
a2	b2	c2	d3
a3	b3	c4	d3

[3]



UNIVERSITY of
ROCHESTER

Normalization

Normalization

- ▶ **Normalization**: The process of decomposing unsatisfactory "bad" relations by breaking up their attributes into smaller relations
- ▶ **Normal form**: Defines conditions to be satisfied by a relation schema, using keys and FDs of a relation.

Normal Form

The *normal form* of a relation refers to the highest normal form condition that it meets, and hence indicates the degree to which it has been normalized.



Normalization

Good design criteria

Two additional properties:

1. The *nonadditive join* or lossless join property, which guarantees that no spurious tuples are generated after decomposition
2. The *dependency preservation* property, which ensures that each functional dependency is preserved after decomposition



Normalization

- ▶ A **superkey** of a relation schema $R = \{A_1, A_2, \dots, A_n\}$ is a set of attributes $S \subseteq R$ with the property that no two tuples t_1 and t_2 in any legal relation state r of R will have $t_1[S] = t_2[S]$
- ▶ A **key** K is a superkey with the additional property that removal of any attribute from K will cause K not to be a superkey any more.

Normalization

Keys

- ▶ If a relation schema has more than one **key**, each is called a **candidate key**.
 - ▶ One of the candidate keys is arbitrarily designated to be the **primary** key, and the others are called **secondary** keys.

Prime attribute

An attribute of relation schema R is called a **prime attribute** of R if it is a member of some candidate key of R.

Nonprime attribute

An attribute is called **nonprime** if it is not a prime attribute that is, if it is not a member of any candidate key.

[4]



UNIVERSITY of
ROCHESTER

Normalization

Keys

EMPLOYEE

Fname	Minit	Lname	<u>Ssn</u>	Bdate	Address	Sex	Salary	Super_ssn	Dno
-------	-------	-------	------------	-------	---------	-----	--------	-----------	-----

DEPARTMENT

Dname	<u>Dnumber</u>	Mgr_ssn	Mgr_start_date
-------	----------------	---------	----------------

DEPT_LOCATIONS

<u>Dnumber</u>	<u>Dlocation</u>
----------------	------------------

PROJECT

Pname	<u>Pnumber</u>	Plocation	Dnum
-------	----------------	-----------	------

WORKS_ON

<u>Essn</u>	<u>Pno</u>	Hours
-------------	------------	-------

DEPENDENT

<u>Essn</u>	<u>Dependent_name</u>	Sex	Bdate	Relationship
-------------	-----------------------	-----	-------	--------------

Figure 3.5

Schema diagram for the COMPANY relational database schema.

Normalization

First Normal Form

- ▶ Disallows attributes whose values for an individual tuple are *non-atomic*
 - ▶ composite attributes
 - ▶ multivalued attributes
 - ▶ nested relations
- ▶ Considered to be part of the definition of a relation
 - ▶ Most RDBMSs allow only those relations to be defined that are in 1NF




Functional Dependencies

First Normal Form

(a)

DEPARTMENT

Dname	<u>Dnumber</u>	Dmgr_ssn	Dlocations



(b)

DEPARTMENT

Dname	<u>Dnumber</u>	Dmgr_ssn	Dlocations
Research	5	333445555	{Bellaire, Sugarland, Houston}
Administration	4	987654321	{Stafford}
Headquarters	1	888665555	{Houston}

DEPT_LOCATIONS

<u>Dnumber</u>	<u>Dlocation</u>
1	Houston
4	Stafford
5	Bellaire
5	Sugarland
5	Houston

(c)

DEPARTMENT

Dname	<u>Dnumber</u>	Dmgr_ssn	<u>Dlocation</u>
Research	5	333445555	Bellaire
Research	5	333445555	Sugarland
Research	5	333445555	Houston
Administration	4	987654321	Stafford
Headquarters	1	888665555	Houston

First Normal Form

Nested Relations

EMP_PROJ

		Projs	
Ssn	Ename	Pnumber	Hours

(b)

EMP_PROJ

Ssn	Ename	Pnumber	Hours
123456789	Smith, John B.	1	32.5
		2	7.5
666884444	Narayan, Ramesh K.	3	40.0
453453453	English, Joyce A.	1	20.0
		2	20.0
333445555	Wong, Franklin T.	2	10.0
		3	10.0
		10	10.0
		20	10.0
999887777	Zelaya, Alicia J.	30	30.0
		10	10.0
987987987	Jabbar, Ahmad V.	10	35.0
		30	5.0
987654321	Wallace, Jennifer S.	30	20.0
		20	15.0
888665555	Borg, James E.	20	NULL

(c)

EMP_PROJ1

Ssn	Ename
-----	-------

EMP_PROJ2

Ssn	Pnumber	Hours
-----	---------	-------

Functional Dependencies

Second Normal Form

- ▶ Uses the concepts of *FDs* and *primary key*
- ▶ Definitions
 - ▶ **Prime attribute**: An attribute that is member of the primary key K
 - ▶ **Full functional dependency**: a FD $Y \rightarrow Z$ where removal of any attribute from Y causes FD to not hold any more
 - ▶ $X \rightarrow Y$ is a **partial dependency** if some attribute can be removed from X and the dependency still holds

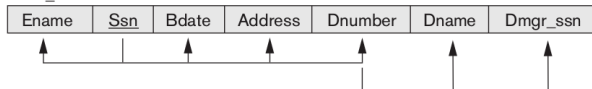


Functional Dependencies

Second Normal Form

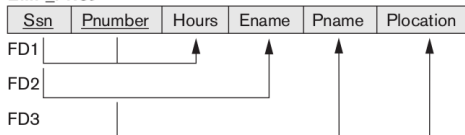
(a)

EMP_DEPT



(b)

EMP_PROJ



[6]



UNIVERSITY of
ROCHESTER

Functional Dependencies

Second Normal Form

- ▶ Uses the concepts of *FDs* and *primary key*
- ▶ Definitions
 - ▶ **Prime attribute**: An attribute that is member of the primary key K
 - ▶ **Full functional dependency**: a FD $Y \rightarrow Z$ where removal of any attribute from Y causes FD to not hold any more
 - ▶ $X \rightarrow Y$ is a **partial dependency** if some attribute can be removed from X and the dependency still holds
- ▶ A relation schema R is in **second normal form** (2NF) if every non-prime attribute A in R is fully functionally dependent on the primary key

[6]



UNIVERSITY of
ROCHESTER

Functional Dependencies

Second Normal Form

(a)

EMP_PROJ

